

FIGURE 1

[illegible]

FIGURE 2

HIV	RT	VKLKPGMDGPKVKQ	WPLTEEKIKALVEICTEMEKEGKISKIGPENFYNTFVFAIKKKDSTKVR	
HTLV1	RT	RPWARTPPKAPKQ	PVPFPERLQALQHLVRKALEAGHIEPTY	PGNNPVFPVKKA NGTVR
Ec-67	RT	NLYRIGSDNQYQTFIT	PKGKGVRTISAPTDRL	KDIQRICDDLSDCRDEIFAIRKI SNNYS
Mx-162	RT	AFHREVDTATHYVSWTIP	KRDCSKRTITSKPEL KAAQR	VWLS NVV ERLP VHCAA
			o o o o o o o o o o x x x o o o	x o
HIV	RT	KLVDFRELNKRTQDFVEV	QLGIPHAGLKKK KSVTVLVDGVDAYFVSP	LDEDFRKYT
HTLV1	RT	FIHDLRATNSLTIDLSSSS	QGPLDSSLPITLAHLQITIDLRAFFQIPLPKQFP	PFY
Ec-67	RT	FGFE	RCKSIILNAYKHRCQIILNIDLKOFFESFNFCGRV	YFLS NQDF
Mx-162	RT	HGCV	ACRSILTNALAHQCADVVVVDLKDFFPSVTVRRVKGLLRKGC	LRECTSILLSSLSTEAP
			o o o o x x x o o o x x x o o o x x	
HIV	RT	FTIP	SINNETPGIRYQYNVLPGCWKGS	PAIFQS SMTKILEPFKKQNPDIVIYQYHDDL
HTLV1	RT	FTVP	QQCYGPGCTRYAKHLVPGCKNS	PLTFEM QLAHLIQPIRQAFPGCTILCYHDDL
Ec-67	RT	LN	PVVATTLAKAACYN GTLPGQSPCSPIISNL	CINHMRLAKLAKKY GCTYSRYADDTI
Mx-162	RT	REAVQF	RGKLLHVAKGP RALPGCAFTSPGITHAL	CLDKRLSALAKRL CFTYTRYADDTF
			o o o o o o o o o o x x x x x x	o o
HIV	RT	S	DLEIGQHRTKIEELRQHLLRWGLTTP	DKKHQKEP PFLWMGYELHPDKVTVQPIVLP
HTLV1	RT	S	PSHEDLLLLSEATHSLISHLGPN	ENKQTPGCTIKFLGQIISPHNLTYDAVPTPI
Ec-67	RT	STNK	TFPLEMATQPEGVVLGKVLKVEISEGFE	INDSKETRLTYKTSRQEV GTLVNRVIND
Mx-162	RT	SWTKAKQPKPR	TQRPPVAVLLSRVQEVVEAEGRVHPDKTRVARKG	TRQRVT GLVVNAAGKDA
			* o o o o o o o o x x x x x x	
HIV	RT	SWTVNDIQIKVLGCKL	MWASQIYP	
HTLV1	RT	RWALPELQALLGEIQW	VSRTGE	
Ec-67	RT	RCYYKKTALAHAL	YRKE YK	
Mx-162	RT	PAARVP	PDVVRQLAAIHN RK	

FIGURE 3

A

Mx-162 18 PTPELTAPSSDAAAKREARRLAHEALLVRAKAIDEAGGADDWVQAQLVSKGLAVEDLD-FSSASEKDKKA-WKEKK 91

Mo-MLV 1070 PDPDMTRVTNSPSLQAHLQALYLVQHEVW-RPL-AAAYQEQ-LDRPVVPHFYRVGDTVWVRHQTKNLEPRWKGPY 1142

Mx 162 92 KAEATERRALKRQAHEAW-KATHVGHLCAGVHWAEDRL 128

Mo-MLV 1143 TVLLTTPTALKVDGIAAWIHAAHVKAADPGGG-PSSRL 1179

B

Mx-162 411 GKDAAPARVPRDVRQLRAAIHNRKKKPGREGESLEQLKGMAAFIMTD-PAKGRAF-LAQLTELESTASAAPQAE 485

HIV 396 GKEGHSARQCR-APR--RQGC--WKCKKPGHIMTNCPPD-R-QAGFLGLGPNKPKPRNFPVAQVPQ-GLTPTAPP 461

Figure 4. Sequence Similarity of the msDNA-Mx162 Reverse Transcriptase with Other Retroelements

(A) Sequence similarity of the region from residues 18 to 128 of the msDNA Mx162 RT (see Figure 2) with a carboxy-terminal region of integration protein of Moloney murine leukemia virus (M-MuLV) (residues 1070 to 1179; Shinnick et al., 1981)

(B) Comparison of the sequence from residues 411 to 485 of the msDNA-Mx162 RT (see Figure 2) with the sequence from residues 396 to 461 of the gag protein of human immunodeficiency virus (HIV; Ratner et al., 1985).

FIGURE 4

A

Mx-162	304	GP-RALPQGAPSPGITHNALCLKDKRLSALAKRL-GFTYTRYADLTF-SWTKAKQPKPRRTQRPVAVL	371
Ec-67	159	YN-GTLPGQSPSPISNLICINDMRLAKLAKKY-CCTYSRYADEITI-STNKWTFPLEMATVQPEGVVL	226
Ec-86	130	YK-NLLPQGAPSSPKLANLICSKLDYRIQGYAGSR-GLIYTRYADLTL-SAQSMKKVVKARDFLFSIIPS	197
HIV	311	YQYNVLPQGWKGSPIAFQS---SMTKILEPFKKQNPDIYIYQYMDLLVGS-DLEIGQHRTKIEELRQHLL	377
HTLV1	150	YANKVLPQGFKNSTLFEH---QLAHILQPIRQAFQCTILQYMDILLAS--PSHEDLLLSEATHASLI	215
Mo-MLV	303	LTWTRLPGGFKNSTLFEH---ALHRDLADFRIQHPDLILLQYVDELLAA-TSELDCCQG-TRALL-QTL	367
RSV	141	FQWKVLPQGMTCSPTICQL---VVGQVLEPLRLKHPSLCMLHYMDLLAA--SSHGGLAAGEEVI-STL	205
BLV	122	FAWRVLPQGFINSALFER---ALQEPLRQVSAAFSQSLVSYMDILYAS--PTEEQRSQCYQALA-ARL	186
Mt. plasmid	288	IATNGVPQGAFTSCGLATYNVL-----ELFLRY--DELIMYADDGIL-CRQDPSTPDFSVEEAGVVQEP	348
17.6	339	YEYLRMPFGLKNAP-ATFORCMN-DI---LRPLLKHC-LVYLDHIVFS-TSLDEHLQSLGLVFE--KL	399
GYPSY	284	YEFCLPFGLRNASSIFQR---ALDDV---LREQI-GKICYVVDVVIIFS--ENESDHVRHIDTVLK-CL	344
Copia	1032	CKLNKAIYGLKQAARCMFR-CIYI---LDKGNINENIYV-LLYVDVVIAT--GDNTRMNNFKRYLME-KF	1112
Tal-3	990	CLLKSLYGLKQSPQWNA-CVYV-KQVSE-QEHLYL---LLYVDHMLIAG--KSKSEINKVKEQLSM-EF	1069
Ty912	948	IRLKSLYELKQS-GANWYE--EVRG-WSCVFKNSQV-TICLFVDDHVLFS--KNLNSNKRRIEKLQM-QY	1023

B

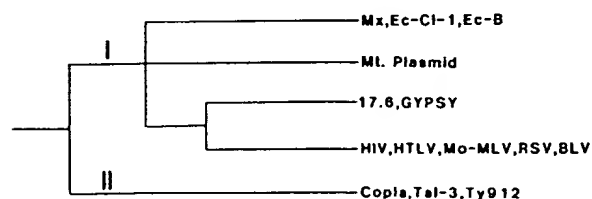


FIGURE 5

S 1 2 3 4



08808031-030397

FIGURE 6

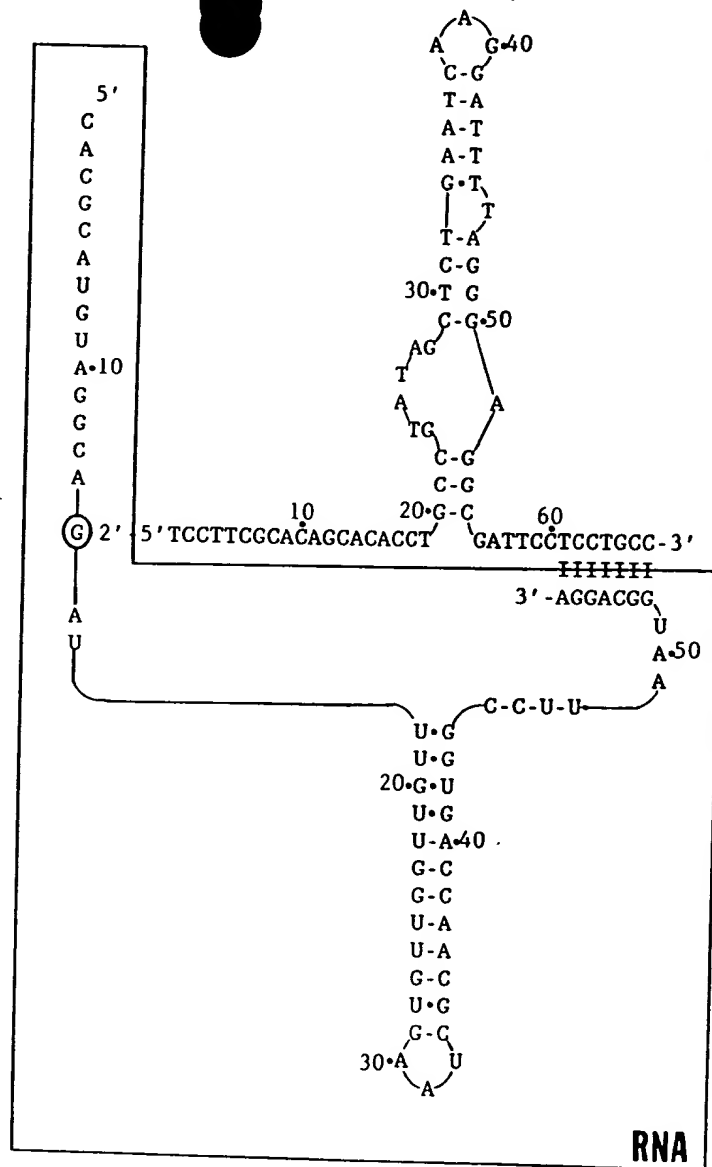


FIGURE 7

030397-00000000

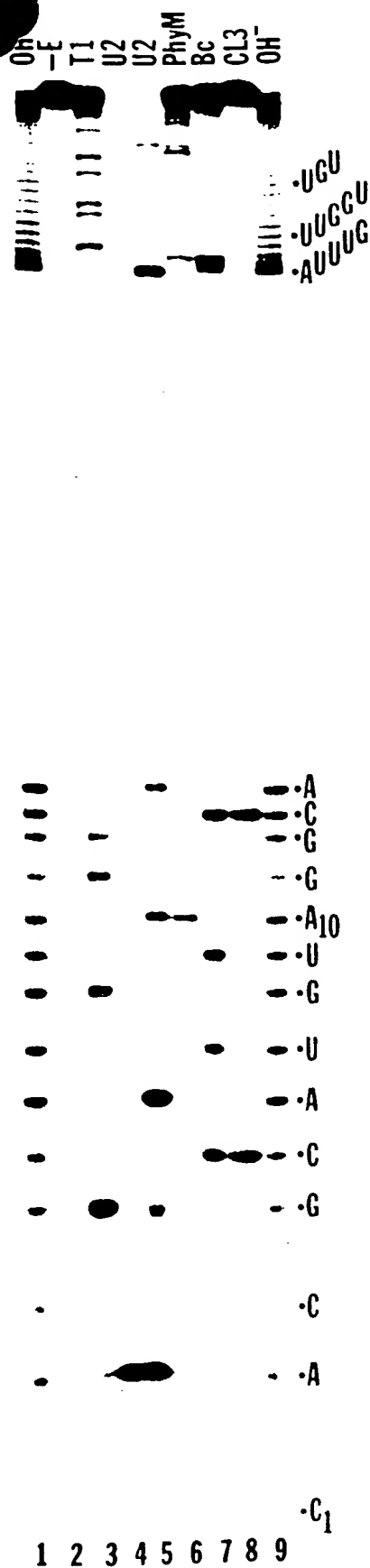


FIGURE 8

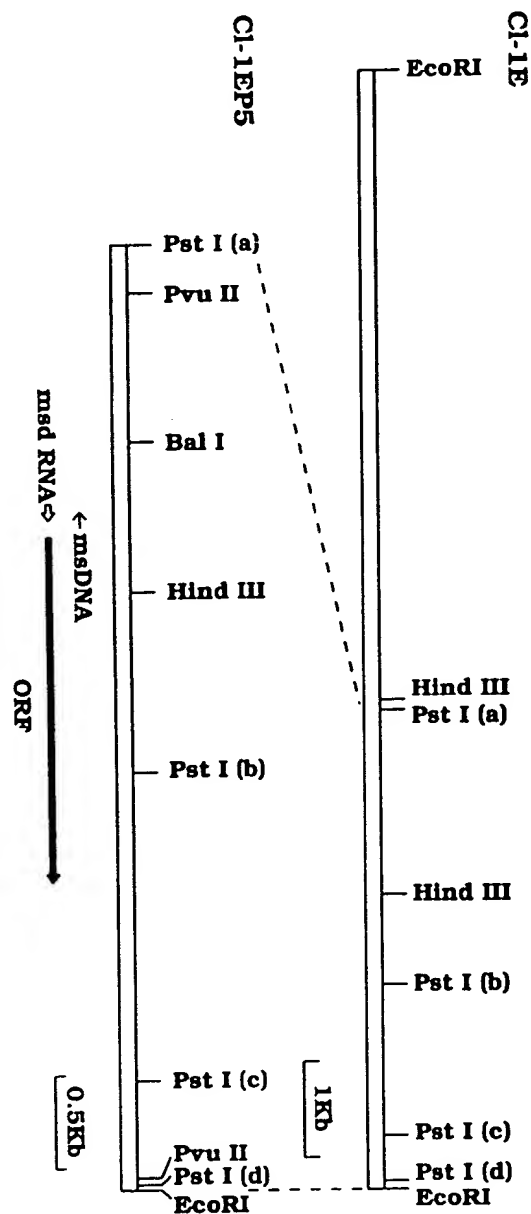


FIGURE 10

08808031-030397

TCG CCA TTT TCA GAT CTT TCA GAT CCG CCA GAA CTC GCG GCG TTT TCC TCA 170
TGT CAT CCA TGT CCA TCA AAA CCA CTC CAT AAA CCG GCG AGC CCG GCG CAT ACC ACC 180
CGC GCG TAT CAC CCA AAA TAG CCA AAA TAC TTC TCG AAA ACA GAA ACT TGA ACT GAT ATG 240
TTC ATA AAC ACC CAT CTA GCG ACC TTT CTT CCG TGT GAA TCG GAA CCA CTC GCG TTA ATG 300
AAG TAT TTC TCG CTA CAT CCG TCT AAA CAA CCA ACA CTT ACC GTT GGT CAC CCG AAT TAG
CGA GCG CCA ATG GCG TCG CTA AAA TCG TTG ATT CAG ACC TAT ACC GGA GGT CTC CTC TCG 360
CGT CCG CCG TAG GCG ACC GAT TTT AGC AAC TAA CTC TCG ATA TCG CCG CCA CAC GAC ACC
GAA GGA CTC CCG CCA TCG CTT TCT CCG CCG TTT TTC CTC TCG CAT CAA GAA CAA ATG 420
CTT CCG CAC CCA CCG ACC CAA AGA CCA ACC GGA AAA AAG CAG ACC CTA CTT CTT CTT TAC
ACA AAA ACA TCT AAA CTT GAC CCA CTT AGC GCT CCG ACT TCA CCG GAA CAC TTC GCT AAA 480
T K T S K L D A L R A A T S R E D L A K
ATT TTA CAT ATT AAG TTG GTA TTT TTA ACT AAC GTT CTA TAT AGA ATC CCG TCG CAT AAT 540
I L D I K L V F L T N V L Y R I C S D N
CAA TAC ACT CAA TTT ACA ATA CCG AAC AAA GGA AAA GCG GTA ACC ACT ATT TCT CCA CCG 600
Q Y T Q F T I P K K G K C V R T I S A P
ACA GAC CCG TTG AAG CAC ATC CAA CCA AGA ATA TGT GAC TTA CTT TCT GAT TGT AGA CAT 660
T D R L K D I Q R R I C D L L S D C R D
GAG ATC TTT GGT ATA ACC AAA ATT AGT AAC AAT TCC TTT CCG TTT GAG ACC GGA AAA 720
E I F A I R K I S N N Y S F G F E R C K
TCA ATA ATC CTA AAT CCG TAT AAG CAT AGA CCG AAA CAA ATA ATA TTA AAT ATA CAT CTT 780
S I I L N A Y K H R G K Q I I L N I D L
AAG GAT TTT TTT CAA ACC TTT AAT TTT CCA CCA CTT ACA GGA TAT TTT CTT TCC AAT CAG 840
K D F F E S F N F G R V R G Y F L S N Q
GAT TTT TTA TTA AAT CCG CTC GTC CCA ACC ACA CTT GCA AAA CCG CCA TCC TAT AAT GCA 900
D F L L N P V V A T T L A K A A C Y N C
ACC CTC CCG CAA CCA AGT CCA TGT TCT CCG ATT ATC TCA AAT CTA ATT TCG AAT ATT ATG 960
T L P Q G S F C S P I I S N L I C N I N
CAT ATC AGA TTA CCG ATC GCG CCA AAA TAT GCA TGT ACT TAT ACC AGA TAT CCG GAT 1020
D H R L A K L A K K Y G C T Y S R A L
GAT ATA AGA ATT TCT ACA AAT AAA AAT AGA TTT CCG TTA CAA ATG CCG ACT CTC CAA CCG 1080
D I T I S T N K N T P P L E H A T V Q P
CAA CCG CCG TTT TTG CCA AAA CTT TTG GTA AAA CAA ATA CAA AAC TCT CCA TTC CAA ATA 1140
E G V V L G K V L V K E I E N S G F E I
AAT GAT TCA AAG ACT AGC CTT AGC TAT AAG ACA TCA ACC CAA CAA CTA ACC GCA CTT ACA 1200
N D S K T R L T Y K T S R Q E V T G L T
GTT AAC AGA ATC GTT AAT ATT CAT AGA TGT TAT TAT AAA AAA ACT CCG CCG TTG CCA CAT 1260
V N R I V N I D R C Y Y K K T R A L A N
CGT TTG TAT CCG ACA CCG CAA TAT AAA CTC CCA CAT AAT CCG GTT TTA GTT TCA GCA 1320
A L Y R T G E Y K V P D E N G V L V S C
CGT CTC GAT AAA CTT CAG CCG ATC TTT CCG TTT ATT CAT CAA CTT CAT AAG TTT AAC AAT 1380
G L D K L E C M F C F I D Q V D K F N N
ATA AAG AAA AAA CTC AAG AAC CAA CCG GAT AGA TAT GTA TTG ACT AAT CCG ACT TTG CAT 1440
I K K K L N K Q P D R Y V L T N A T L N
CGT TTT AAA TTA AAG TTG AAT CCG CCA GAA AAA CCA TAT ACT AAA TTT ATT TAC TAT AAA 1500
G F K L K L N A R E K A Y S K F I Y Y K
TTT TTT CAT CCG AAC ACC CCG CCG ACC ATA ATT ACA CAA CCG AAC ACT CAT CCG ATA TAT 1560
F F H G N T C P T I I T E C K T D R I Y
TTG AAG CCG CCG TCG CAT TCT TTG GAC ACA TCA TAT CCG CAG TTG TTT AGA GAA AAA ACA 1620
L K A A L H S L E T S Y P E L F R E K T
GAT AGT AAA AAG AAA CAA ATA AAT CTT AAT ATA TTT AAA TCT AAT CAA AAG ACC AAA TAT 1680
D S K K K E I N L N I F K S N E K T K Y
TTT TTA CAT CTT TCT CCG CCA ACT CCA CCG CAT CTC AAA AAA TTT GTA CAG CCG TAT AAA AAT 1740
F L D L S G C T A D L K K F V E R Y K N
AAT TAT CCG TCT TAT TAT CCG CTT CCA AAA CAG CCA GTC ATT ATG CTT CTT GAT AAT 1800
N Y A S Y Y G S V P K Q P V I H V L D N
GAT ACA CCG CCA ACC GAT TTA CTT AAT TTT CTC CCG AAT AAA CTT AAA ACC TCG CCA CAC 1860
D T G P S D L L N F L R N K V K S C P D
CAT GTA ACT GAA ATC AGA AAG ATC AAA TAT ATT CAT CTT TTC TAT AAT TTA TAT ATA CTT 1920
D V T E H R K N K Y I H V F Y N L Y I V
CTC ACA CCA TTG ACT CCG TCG CCG CAA CAA ACT TCA ATC GAG CAT CTT TTC CCG AAA GAT 1980
L T P L S P S G E Q T S H E D L F P K D
ATT TTA CAT ATC AAG ATT CAT CCG AAC AAA TTC AAC AAA AAT AAT GAT CCA CAC TCA AAA 2040
I L D I K I D G K K F N K N N D G D S K
AGC GAA TAT CCG AAG CAT ATT TTT TCG ATC ACC CTT CTT ACA GAT AAA AAG CCG AAA ATA 2100
T E Y G K H I F S M R V V R D K K R K I
GAT TTT AAG CCA TTT TGT TGT ATT TTT GAT CCG ATA AAA GAT ATA AAG GAA CAT TAT AAA 2160
D F K A P C C I F D A I K D I K E H Y K
TTA ATC TTA AAT ACC TAA TCA ACA CCG CTA ACC TTA TCA ACC CTA ACC CTC ATT TTT CCG 2220
L H L N S
TAA AAT TTA TAT CCG TTG AAT TGT AAT ATA TTA TCT TCA ACC CAT TTA TTT AAT TCG TCG 2280
ATC CTT TTC TGT AAC CCG ATT AAT TCG TTC CTC ACA AAG ACT AAA CTC CCG TTT TCG ACA 2340
TCC CCA AAG CCG CCG AAT ATT ATT CCG CAT AAT CCG CAT CAT TTC CCG TCG CAC ACC ATC 2400
CGC TCG CAT CAT CTC ATC CCG CC

FIGURE 11

V	RT	VKLKPGMDGPKVKQ	WPLTEEKIKALVEICTEMEKEGKISKIGPENPYNTPVFAIKKKDSTKWR	239												
LV1	RT	RPWARTPPKAPRNQ	PVPFKPERLQALQHLVRKALEAGHIEPYTG	PGNNPVFPVKKA	NGTWR	75										
DNA	RT	NVLYRIGSDNQYTQFTIPKKGKGVRTISAPDRL	KDIQRRICDLLSDCRDEIFAIRKI	SNNYS	94											
		+ ○	● ● ○	+	●++ ● +											
V	RT	KLVDFRELNKRTQDFWEVQLGIPHPAGLKKK	KSVTVLDVGDAYFSVPLDEDFRKYTAFTIP	SI	302											
LV1	RT	FIHDLRATNSLTIDLSSSSPGPPDLSSLPTTLAHLQ	TIDLRLDAFFQIPLPKQFQPYFAFTVP	QQ	139											
DNA	RT	FGFE	RGKSIILNAYKHRGKQIILNIDLKDFEF	SFNFGRVRG	YFLS	NQDF	LLN	PVVA	150							
		○ ●	+	●+	+	+○	●+	●+	●							
V	RT	NNETPGIRYQYNVLPQGWKGSPAIFQS	SMTKILEPFFKKQNPDIVIYQ	VMDDLYVGS	DLEIG	363										
LV1	RT	CNYGPGTRYAWKVLPQGFKNSTPLFEM	QLAHILQPIRQAFPOCTILO	VMDDILLAS	PSHE	199										
DNA	RT	TTLAKAACYN	GTLPGQSPCSP	PIISNLICNIMDMRLAKLAKKY	GCTYSR	VADDITI	STNKNTF	212								
		●	●●●●	●●+	+	○	+	○○	●●●●	●						
V	RT	QHRTKIEELRQHLLRWGLTTP	DKKHQKEP	PFLWMGYELHPDKWTVQPIVLPE	KDSWTVNDI	424										
LV1	RT	DLLLSEATMASLISHGLPVS	ENKTQQTPTGIKFLGQIISP	NHLYDAVPTVPI	RSRWALPEL	262										
DNA	RT	PLEMATVQPEGVVLGKVLVKEIENS	SGFEINDSKTRLTYKTSRQEV	T	GLTVNRIVNIDRCYKKT	276										
		○	+	●	○○	○+	○	●	+	+						
V	RT	QKLVGKLNWASQIYPGIK	VRQLCKLLRGTKALTEVIPLTEEA	ELELAENREILKEPVHGVYYD	487											
LV1	RT	QALLGEIQWVSKGTPTLRQPLHS	LYCALQRHTDPRDQIYLNPSQVQSLVQ	LRQALSQNCRSRLVQ	327											
DNA	RT	RALAHALYRTGE	YKVPDE	NGV	LVSGGLDKLEG	MFGFIDQVDKFN	NIKKLKNQ	PDRYVL	335							
		○○	+	+	+	+	○	+	○○	○+						
V	RT	PSKDLIA	EIQKQGQGWTYQIYQE	PFKNLKTGKYARMRGAHTNDVKQLTEAVQKITT	544											
LV1	RT	TLPLLGAIMLTLTGTTTVFQSKEQWPLVWLHAPLPH	TSQCPWGQLLASAVLLLDKYTLQSY	GL	391											
DNA	RT	TNATLHGFKLKL	NAREKAY	SKFIY	YKFFHGNTCPTIITEGKTDR	IYLKAALHSLET	SYPEL	396								
		○	●	○ ○	+	○○	+	○	○	+	+	+	○	+	○○	○
V	RT	ESIVIWGKTPKFKLPIQKETWETWWTEYWQATWI	PE	WEFV	NTPPL	VKLWYQ	595									
LV1	RT	LCQTIHNNISTQTFNQFIQTS	DHPSVPILLHSHR	FKNLGAQTGELWNTFLKTAAPLAPVKALMP	456											
DNA	RT	FREKTD	SKKKEINLNIFKSNEKTYFLDL	SGGTADLKKFVERYKNNYASYYGSV	PKQP	VIMVLD	460									
		+	+	○ ○	+	○	○	●	○○							
V	RT	LE	KEPIV	GAETFYVDGAANRETKLGKAGYVTNKG	RQK	VV	PLTNTTNQ	KTELQAIYLA	652							
LV1	RT	VFTLSP	VIINTAPCLFSDG	STSRAAYILWDKQILS	QRS	FP	LPPPHKSA	Q	RAELLGLLHGL	516						
DNA	RT	NDTG	PSDLLN	FLRNKVKSCPD	DVTEMRKMKYIHVFYNLYIVLTPLSP	SGEQTSMEDLFPKDIL	523									
		○	●	○	+	+	+	○+	+	○	●	○	●	○	○	
V	RT	LQDS	GLE	VNIVTDSQYAL	QIIQA	QPKDSESELVNQIIEQLIKKEKVYLAWVPAHKG	708									
LV1	RT	SSAR	SWR	CLNIFLDSKYLYHYLRTLALGTFQGRSSQAPFQA	LLPRLLSRKVVYLHVRSH	TN	578									
DNA	RT	DIKIDGKKFNKNNDGDSKTEYGKHI	FSMR	VV	RDKKRKIDFKAFCCIFDA	572										
		+	●	●●○	○	+	○	○	○	+	+	+	+	+	+	
V	RT	IGGNEQVDKLVSAG														722
LV1	RT	LPDPISRLNALTDA														592
DNA	RT	IKDIKEHYKLMLNS														586
		+	○	++												

FIGURE 12

00000001-000000

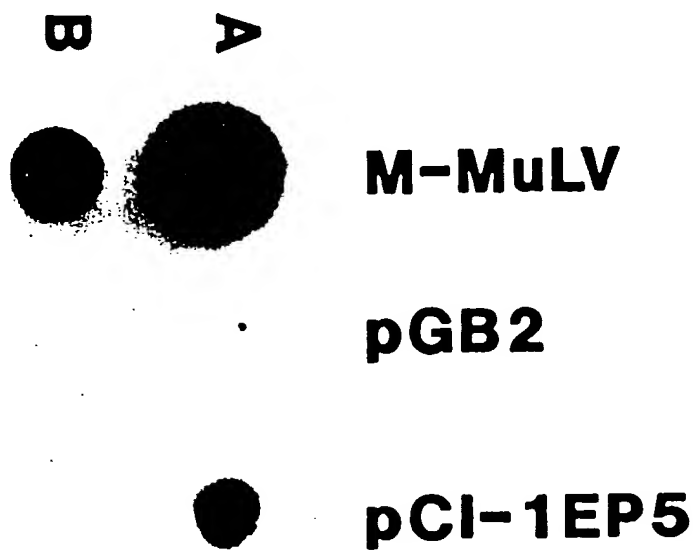


FIGURE 13

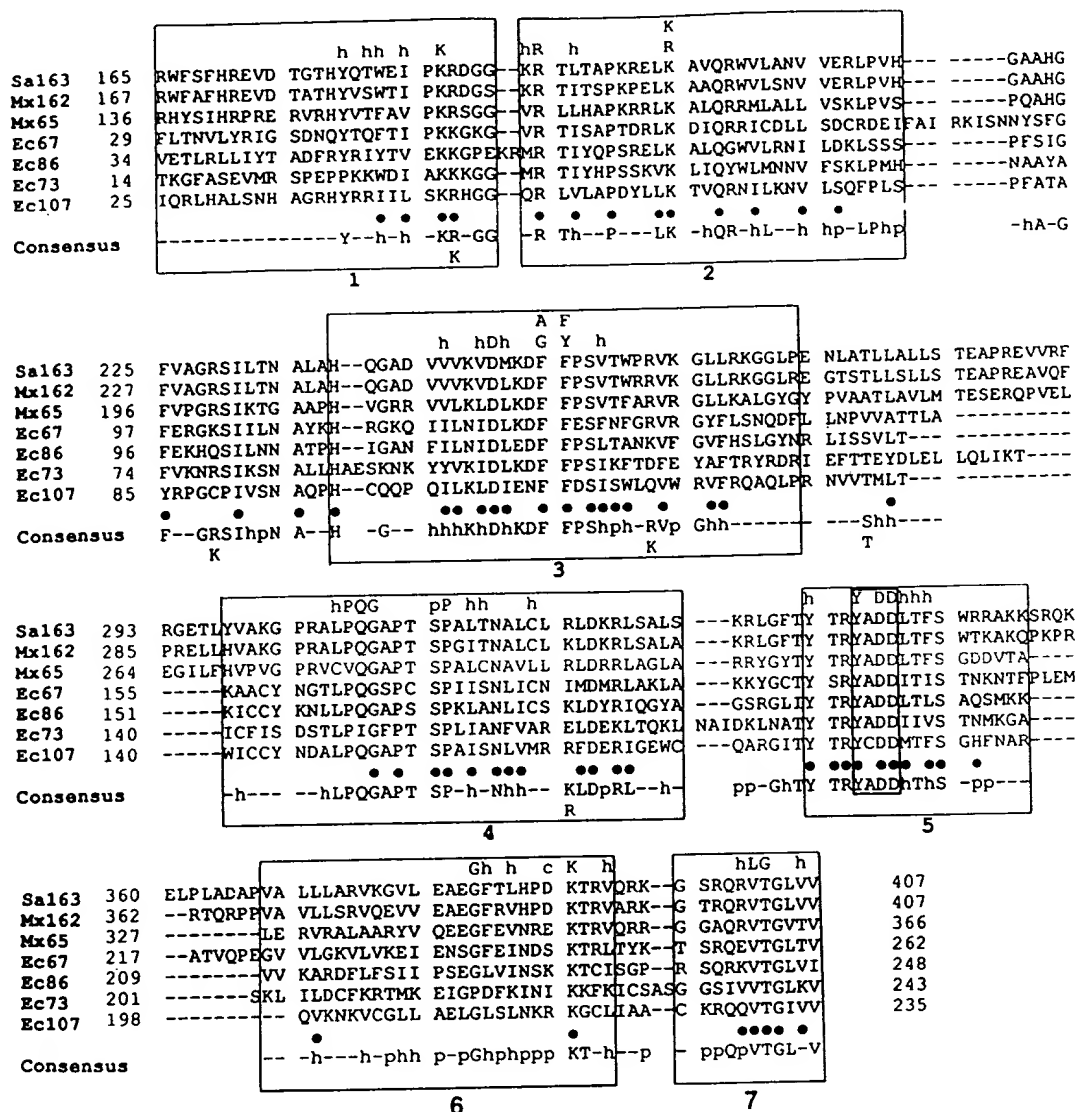


FIGURE 14

CC ACT TCC GGC GCT CGG GCT GCG GGA GGG CCC GTG GGA GCA CAT GAT GGC GCT GCG GCT 60
 GT CCA GGT CGG GCA CGC CGC GGA GCA GGA AGC ACT GCG TCA GAC CCC CGC GGG CGG CCA 120
 CT CAT CGG CGC GGA GAC CGC CTC CTA GGT GCG GCG GGA GGC CTC CGG CCA GGA GCA GGT 180
 TA CGG GGT CTC ATT GGA TGG GAA AGT GGT GCG GGT GGA GTG GGG CGC CGC CCA GGG GGA 240
 TC CGG CGG GGA GAA GGT CTC GTT GGA CAC GGA CGC GGA GGC CGC CAC CGC CTA CTT CAC 300
 TG CTT GGA GTC CTT GCG CGC GGA GGG ATA TAT GGA TGC GGC TGC TTC AAT GAT GTA GAA 360
 AC GAA AGC CAC GCG CGC CGC GCG CGG AAA CGC AGC TGC CAC GGA AGC ACA GAC 420
 31 RNA< 1
 TT CGT GCG AGC GAC CCA GAG AGC TCC CAA GGC ATC AGC CTC AGC CAC TCC AGC CGC AGA 480
 TA GCA CGC TCG CTC GCT CTC TCC AGC GTT CGG TAG TCG GAG TCG CGC AGC TCG CGC TCT 540
 TG GCG TTG CGC CGC TCT GGT TGA ATT GGA GGA CAC TCT CGG CAA GGT AGC CTG TTC TTG 600
 TC CGC AAC CGC CGC AGA CCA ACT TAA GGT CTT GAT AGA GGC GTT CCA TCG GAC AAG AAG 660
 T CTC TTC CTT CGC GTG AGT ACC TCT CGG CGC GCG GAG CTG AAC CAA GGA CGC AAC CGC 720
 A GAG AAG GGA GGC CAC TCA TGG AGA GGC CGC CGC CTT GAG TTG GTT GCT GCG TTG GCG 780
 TT TTC CGC CGC CGC AGA GGT ACT CAC CGC AGC GGA GAG CGC GTG AGC CTA CGC TCC CGC 840
 A AAG GCG CGC CGC TCT CCA TGA GTG GGC TCC CTT CTC GGC CAC TCC GAT GCG AGC GCG 900
 31
 TG TGA GAA GGT GGT GGC TTC GCG CTT CGC TCG AGC GCT CGC GCT CGC TCG CGC TCG CCA 960
 TC AGT CTT GGA CCA CGC AAG CGC GGA CGC AGC TCG CCA GCG CGA GCG CGC AGC AGC CGT 1020
 TC TCG CGC CGC CCA CTT TCG TCA CGG GCG CCA GGA GGC GTC ATG ACC GCG AAG CTG GAG 1080
 CA CAC GTC CGC GCG GCG CGC CGC GTC TCG GCG GCG GCG CGC CGC CGC CGC CGC CGT CGC GAT 1140
 TC GCG AAG CAG GAG GCG CGC CGC GCG CAC CAC GAG GCG CTG CGC CTG CGC TCG AAG CGC 1200
 TC GAA GAG GCG GCG GCG AGC GAC CGC TCG GTG CGG CAG CAG CTG GTG GCG AAG GCG GTC 1260
 TC GCG GCG GAG GAG GAG TTC GAG TCG CTC AGC AAG CAG AAG GCG GCG TCG AAG GAG 1320
 AG AAG AAG GCG GCG GCG ACC GAG CGC CGC GCG AAG CGC CTG GCG TCG GAG GCG TCG 1380
 K K K A E A T E R R A Q K R L A W E A W 1440
 K GGC AAT CAC ATC CAC CTG GCG GTG GCG GTG CAC TCG GAG GAG CGC GCG GCG GCG 1500
 K A H I H H L G V G V H W D E A G G G P 1560
 AC AAG TTC GAC GTG GCG GCG GCG GAG GAG CGC GCG AAG GCG AAG GCG TTG CGC GAG GCG 1620
 D K F D V A G R E E R A K A H G L P E G 1680
 TG GAC TCG GTC GAG GCG CTG GCG AAA GCG CTG GCG ATC TCC GTG TCG CGC CTG CGC TCG 1740
 L D E V E A L A K A L G I S V S R L R W 1800
 TC TCC TTC CAC CGC GCG GTG GAC AGC GCG AGC CAC TAC CAG AGC TCG GAG ATT CGC AAG 1860
 F S F H R E V D T G T H Y Q T W E I P K 1920
 CG GAC GCG GCG AAG CGC AGC CTC ACC CGC CGC AAG CGC GAG CTC AAG GCG GTG CAG CGC 1980
 R D G G K R T L T A P K R E L K A V Q R 2040
 CG GTG CTC GCG AAC CTG GTG GAG CGC CTG CGC GCG GCG GCG GCG GCG GCG GCG GCG GCG 2100
 W V A H V V E R L P V H G A A H G F V 2160
 4200
 CG GTG CTC GCG AAC CTG GTG GAG CGC CTG CGC GTG CAC GCG GCG CAC GCG TTC GTG 2220
 M V A H V V E R L P V H G A A H G F V 2280
 CAC GAA GAA GAG CGC ATC GCG GAC CGC GGT GAG CGC GCG GCG GCG GCG GCG GCG GCG 2340
 A G R S I L T H A L A H Q C G A D V V V K 2400
 GTG GAC ATG AAG GAC TTC TTC CTT TCC GTG AGC TCG CGC CGC GTC AAG GGA CTG CTG CGC 2460
 V D N K D F F P S V T W P R V K G L L R 2520
 250
 AAG GGA GGA CTC CGC GAG AAC CTG GCG AGC CTC CTG GCG CTG TCC ACC GAG GCG CGC 2580
 K G C L P E H L A T L L A L L S T E A P 2640
 CGC GAG GTG GTG CGG TTC CGG GGA GAG AGC CTG TAC GTG GCG AAG GCG CTT CGC CGC CTG 2700
 R E V V R F R G E T L Y V A K G F R A L 2760
 CGC CAG GCG GCG CGC ACC TCT CGC GCG CTG AGC AAC CGC CTG TCG CGC CTG GAG AAG 2820
 P Q G A P T S P A L T H A L C L R L D K 2880
 300
 CGG CTC TCG CGC CTG TCG AAG CGC CTG GCG TTC AGC TAC AGC CGC TAT GCG GAT CAC CTG 2940
 R L S A L S K R L G F T Y T R V A D D L 3000
 ACG TTC TCG TCG CGC CGC AAG AAG TCC CGC CAG AAG GAA CTC CGC CTG GCG GAT GCG 3060
 T F S M R R A K K S R Q K E L P L A D A 3120
 350
 CGG GTG GCG CTG CTC CTG GCG CGG GTG AAG GGT GTG CTG GAG GCG GAG GGT TTC AGC CTG 3180
 P V A L L L A R V K G V L E A E G F T L 3240
 CAC CGC GAC AAG AGC CGC GTG CAG CGC AAG GCG AGC CGC GAG GTG ACG GCG CTC GTG 3300
 H P D K T R V Q R K G S R Q K E L P L A D A 3360
 CTG AAC GAG GCG CGC GAG GCG GTT CGC GGT GCG CGC GTG CGC CGC GAT GTG GTG CGC CGC 3420
 V M E A P E G V P G A R V R D V V R R 3480
 CTG CGC GCG GCG ATC CAC AAC CGC GAG CAG GCG AAG CGC GCG CGC CGC CGC CGC CGC 3540
 L R A A I H M R E Q G K P G P T G E T L 3600
 GAG CAG CTC AAG GCG CTC GCG GCG TTC CTT CAC ATG AGC GAG GCG AAG GCG CGC GCG 3660
 E Q L K G L A A F L H M T D A E K G R A 3720
 450
 TTC CTG CGA CGC CTG GAG GCG CTC GAG AAG CGC CAG ACC GCG TGA CGC TCA CTG GTC GTC 3780
 F L R R L E A L E K R Q T A - 3840
 CGG GCG ATC GCA GCG GCG GCG AGC GAC CGT CAC CGC CCA GAT CTC CAT GCG ATC CTG 3900
 GGG ATT CTG GCG GGT GAA GAA GAC TTC CCA GCG GAG AGC GAG GCG GCG GCG GCG GCG 3960
 TGA CTC CTC CGC CGC GCG GAT CTC CGC GAG GCG CAC CGT TCC GAG GTC GGT GCG ATT GCT 4020
 CAC CCA GCG GTC CGC GCG CCA GCG GCG GCG GCG GCG GCG GCG GCG GCG GCG GCG GCG 4080
 GCG CGT CAG CTT CTC CGC GCA GCG ATC CTC GCG GCG GCG GCG CAA ATC CTT CAG CAG CAG 4140
 GGT GCG CTT GCG GGT GCG ATC GCT GGA CCA CAG CTC CGC GCG GTG GAG GCT GTC ACT CGC 4200
 GCG GAA GTA GAG CAT CGC ATT CAG CGC CTT GAT GCG GCT GCG GCG CCA GCT GTC CGC ACC 4260
 CGG CCA GAT GTC CTT CAC CGC GAC CGT GCG ATG CGA CGT GCG ATC GCT GAC CCA CAG CTC 4320
 CTC GCG CTC GCG CTG GCG CCA GAA CTC GCG CTC GCG TCC CGC GCG GCT GAA GAA GAT CTT 4380
 CGC CGC GAG CGC GGT GAG ATC ATG CGG ATA GAG GCG GCG GAA GAA GCG CAG CTG CTC GGA 4440
 GAG GGT GCG TCT GGA GCA CCA CAG GGT GCG CTC GCG TTC GTG ATT GTC GAG CAG GAA GAA 4500
 GAG CAC GCA GTC CGC CGC GGT GAA CGC GGA GAG GAA GTT GTC CTC GCG CGC GGT GAA GAC 4560
 AGA GGT GGT GGT GGA CAG CGC CAG GGT GCG CCA GAT GAA CAC CTC GTC ATT GAC GTT GCG 4620
 CAC GAA GAA GAG CGC ATC GCG GAC CGC GGT GAG CGC GCG GCG GCG GCG GCG GCG GCG 4680

FIGURE 16

TTTCCAGAAC CGCCATACCA AACAGGGGAT ACAGACCAAC CTGAGGCTGA AAGAGGAAG CTACGGGAC TGGCTGCCA ACTGCCAGCA 9980
 F E K R N Y K Q G I Q T M L Y L K E E S Y G D W L P K C D D
 CCCCAGCA ACAATACCTC ACTCAGACCG GCACAGCCCG GTCTTTTCTT TTCTGGCCAT TCCGACAACT TGAAGAATCC ACTGTTTCA 10080
 P A A T *
 CTTCACCTT TATTCACTT TATCACTAT CAATTTATTA ATAAAAAACC AGAGCTCAAC AGCTGCACTA GTAAAACTG AAAAACTTT 10170
 TATCACCTCC CGCCATGCC CGACTGGACA GATCCAGAAC GAGCAAAAAT CACAAAGGTG ACCAGCTGAC TCTTCACTCT TCACCAACTC 10260
 ATCACCACCT AACCCATGCA TATAAAATGA TAAATATCG AGCTGATCAG TTAATATGCA AAAAACTTT TCTCAGCTCT TCGATAAAG 10350
 AAAATTAATT CACATCAATA GCTTTCTCTT TCAATCTCTT TCAGCTTTAT TCAGAGCTAA TACAGCCAAA CTTTCAATTT TATGCTTTAT 10440
 TTTTAAATTA GTGAGTTAT CCAAGGAGA ACTTAGGAGA ACTCCAAATA CTCTGGATT CTCTGGTTT GGATCTTAAA ATACCCAAAT
 TAGCCATCG CCAATGAGTC ATGCTTTTCC CTAGTATTTT AGCTTTGCCC GTCTTTGAGT TCGCTGAGCG CGCCCTGGGG CCAACCGATC 10530
 ATCGGCTAGC GGTACTCAG TACCAAGCG GATCATAAAA TCGATACCGG CAGCAAGTCA AGCCACTGCG CGCCGACCCC CGCTGCTAG
 AGCCAACTGA TCGAGCTGCT CAATGAGTTT TCGCTTTTCT AGCTCTCTAC CATCAAGGTG CATCAAGGATA TTCTGATGC TGACTCAGCT 10620
 TCGTTTGAAT AGCTGACCA GTTATCTCAA ACCGAGAAAA TCAGGAGATG GTAGTTCCAC GTATTCTAT AAGAGCTAGC ACTGATCGA
 or(116 N L T Q L
 AAAAAAAT GGTACTGAGG TATCTAGAGC AACCCGCTTA TTTTCATCAT TCGTCAAAA JAACAAGTA AAATCTCTG GTAATGAAA 10710
 K K N G T E V B R A T A L F S S F V E K F K V K C P G H V K
 AAAATCTGCT TTCTGCTGTG GTGCTAACAA AACAATGGA GAACCATCAG CAAGACGATT GGAATTAATA AATTTTCTG AAAGGTATTT 10800
 K F V P L C G A N K N N G E P S A R R L E L I N F S E R J L
 GAATAACTGT CACTTTTTCT TTGCTGAAT AGTTTCAAA GAATTAAGCA CCGATGAAGA ATCAATTATC GATAATTAT TAGATATCGA 10890
 M N C H F F L A E L V F K E L S T D E E S L S D N L L D I E
 AGCTGACTTA TCTAAATAG CTGATCATAT TATCATTTGT TTAGAAAGTT ATTCACTCTT CACGGAAGTT GGTGCTTGG CATACAGCAA 10980
 A D L S K L A D N I I I V L E S Y S S F T E L G A F A Y S E
 GCAATTACCG AAGAAATTA TAATAGTTAA CAATACAAA TTTATAAATG AGAAATCAT TATAATATG GGACCAATAA AGGCTATTAC 11070
 Q L R K K L I I V M N T E F I N E K S F I N N G P I K A I T
 TCAGCACTGA CAACAATCTG GTCAATTCTT ACATTATAAA ATGACAGAAG GTATTGAAG TATAGAGCCG TCTGATGGA TTGGGAAAT 11160
 Q Q S Q Q S G H F L N Y K N T E G I E S I E R S D G I G E I
 ATTCGACCCC CTATATGATA TTCTTTCTAA GAACGACAGA GCAATTTCAA GAATTTAAA AAAGAAGAG TTAGATCTCT CCAGTAAGTT 11250
 F D P L Y D I L S K M D R A I S R T L K K E E L D P S N F
 CAATAAAGC TCAGTACGAT TATTCATGA CGTAATTTT GTATCTGCTC CTTCGCACT TAATGAAGT ATCGAAATTA TCACAAAAT 11340
 N K D S V R F I N D V I F V C G P L Q L N E L I E I I T K I
 ATTGCGACA GAAAGCATT ACAAAAAA TCTTCTAAG CACTTGGTA TTCTAATAGC TATTGAATA ATATCATGCA CAATGGGAT 11430
 F G T E S H Y K K M L L E N L G I L I A I R I I S C T M G I
 TTATTATCT TTGTATAAG AATATTATTT TAAATATGAC TTGACATG ACAACATATC ATCAATGTTT AAAGTTTTTT TCTCAAGAAA 11520
 Y V S L Y K Z Y Y F K Y D F D I D M I S S N F K V F F L E N
 CAAGCCAGAA AGGATGAGG TATATGAGAA TATATAGGCT AATTGATTCT CAGACATTGA TGACTAAGGG ATTGCTTCT GAAGTAATGC 11610
 K P E R M R V E N I *
 RT M R I Y S L I D S Q T L N T K G F A S E V N
 GATCACCTGA CCCCCAAAA AAATGGGATA TAGCTAAGAA AAAAGGAGT ATGAGAACAA TTTATCACCC GTCAATCAAA GTTAATTA 11700
 R S P E P P K K M D I A K K K G S M R T I Y N P S S K V K L
 TTCAATATG GTTAATGAAT AATGTTTTT CGAAGCTGCC AATGCATAAT GCTGCATAGC CATTGTTAA AAACCGATCA ATAAAAAGCA 11790
 I Q Y W L N M N V F S K L P N H M A A Y A F V K M R S I K S
 ATGCTTTAT ACATGCCGAA TCAAGAATA AGTATTATCT CAAAATAGAT CTCAAGATT TTTTCCCTTC AATAAAATTT ACTGATTTTG 11880
 N A L L N A E S K N K Y Y V K I D L K D F F P S I K F T D F
 AGTACGATT CACTGTTAT CGAGATGCA TTGAATTTAC TACAGAATAT GATAAGGAGT TACTACAATC TATAAAAAAG ATCTGCTTTA 11970
 E Y A F T R Y R D R I E F T T E Y D K E L L Q L I K T I C F
 TATCAGATAG CACTCTGCT ATCGGCTTTC CTACATCTCC ATTAATTGCA AACTTTGCG CAAGAGAACT TGATGAAAAA CTGACGAAAA 12060
 I S D S T L P I G F P T S P L I A N F V A R E L D E K L T Q
 AACTAAATG AATTGATAAA CTTAATGCA CTATACAGC ATATGCTGAT GATATTATG TCTCTACAAA TATGAAGGG GCTAGCAAT 12150
 K L M A I D K L N A T Y T R Y A D D I I V S T N M K G A S E
 TAATTCTGGA TTGTTTTAAA AGAACAATGA AAGAGATTGG TCCAGACTTT AUAATTAACA TTAATAAAT TAAGATTCT AGTCTCTCG 12240
 L I L D C F K R T N K E I G P D F K I N I K K P K I C S A S
 GAGGAAGTAT AGTAGTTACC GGATTGAAAG TTTGCCAGCA TTTTCATATT ACATTACATA GATCAATGAA AGATAAATA AGATTGATC 12330
 G G S I V V T G L K V C H D F H I T L N R S N K D K I R L N
 TTCTCTTTT ATCAAGGGC ATATTAAAA ATGAAGATCA TAATAAATCT TCTGTTATA TTGCTTATGC AAAAGATATA GACCTCTATT 12420
 L S L L S K O I L K D E D N N K L S G V I A Y A K D I D P H
 TTTATACAAA ACTGAACAGA AATATTATTT AAGAAATTA ATGATTGAC AATCTCCACA ACAGAGTTGA ATAACTTTA TATTTTGGAT 12510
 F Y T K L N R K I Y F Q E I X W I Q K L N M K V E *
 GCACCCCAAT AACTTCATT ATTAATATGG GAACAATATA GCTTTTCAG GATGACCTAC ACTCTAGAGA ATGTGTATAC AAAAGTGTAT 12600
 AAGTTATTTT CAACCTATA TAAATACAG CAATATCAAT GCAATGGCGG CATTTTACG CTCTGTGAT CTTCGCGCAA AATGCTG 12688

FIGURE 17

P HH B X F
V V V V V
400 bp

FIGURE 18

0806031 030527

Oligo 2337
 tcaccctgaaagacctgattgcttacctggaagagaagccggaatgcggaacatctgg 60
 cggcggttaaggcctatcgcaagagttcggcggtttaaaAATATGCGCTGTCAGGCTTT 120
 TTGCTGTGCGCAACGCTGATGCGCTTCAAGATATCGTGTAAATCTGCTTTCGCCAGCAGTG 180
 AACGACACCGCTCGCACTACGCGAAGTTCATAGCACAATTAGACGAAAGCGGTCGTCAC
 GCAATAGCGTTTCCGGCCTTTGTGCGGGAGGCTCGCGAGTCGCTGACTTAACGCCAG 240
 CGTTATCGCAAAGGCCGAAACACGGCCCTCCAGCCGCTCAGCGACTGAATTGCGGTC
 TAGTATGTCCATATACCCAAAGTCGCTTCATTGTACCTGAGTACGCTTCGCGTACGTCGC 300
 ATCATACAGGTATATGGGTTTCAGCGAAGTAACATGGACTCATGCGAAGCGCATGCAGCG
 GCTGACGCGCTCAGTACAGTTACGCGCCTTCGGGATGGTTTAAATGGTATTGCCGCTGTTG 360
 CGACTGCGCGAGTCATGTCAATGCGCGGAAGCCCTACCAAATTACATAACGGCGACAAC
 GCGCCTCTTTTGGCCGCGGTGATGTGGAGAGTGAATGGATGCTACCCGGACAACCCCTTC 420
 M D A T R T T L L
 TGGCGCTCGATTGTTCGGCTCGCCGGGCTGGAGCGCGGATAAAGAAATACAGCGACTGC 480
 A L D L F G S P G W S A D K E I Q R L H
 ATGCGCTCAGTAATCATGCCGACGCCATTACCGACGCTTATCTTTCTAAACGCCACG 540
 A L S N H A G R H Y R R I I L S K R H G
 GTGGTCAGCGCTGGTGTAGCCCTGATTACTTGCTCAAAACCGTACAGCGCAACATTC 600
 G Q R L V L A P D Y L L K T V Q R N I L
 TTAAGAAGCTCTTTTCAATTTCCGCTTTCCCTTTTGTCTACAGCTACCGACGAGTT 660
 K N V L S Q F P L S P F A T A Y R P G C
 GCCCAATCGTCAGCAACCGCGAGCCACACTGCCAACAGCCGAGATCCTGAAACTCGATA 720
 P I V S N A Q P H C Q Q P Q I L K L D I
 TCGAAAACCTTTTTCGATAGCATTAGCTGGTTACAGGTCTGGCGTGTGTTTCGCCAGGCC 780
 E N F F D S I S W L Q V W R V F R Q A Q
 AGTTGCCACGTAATGTGGTAACCATGCTGACCTGGATTGTTGTTATAACGACGCTTAC 840
 L P R N V V T M L T W I C C Y N D A L P
 CGCAGGGGGCACCAACTTCGCCAGCCATTTCGAATCTTGATGCGCCGTTTGTATGAAC 900
 Q G A P T S P A I S N L V M R R F D E R
 GCATAGGGGAATGGTGTGAGGCTCGGGGAATTACCTACACCCGCTACTGCGATGACATGA 960
 I G E W C Q A R G I T Y T R Y C D D M T
 CCTTTTCAGGTCACCTTCAATGCCCGCAGGTAAAAATAAAGTGTGCGGATTGTTAGCGG 1020
 F S G H F N A R Q V K N K V C G L L A E
 AGCTGGGCCTGAGCCTCAATAAACGCAAAGGCTGCCTGATAGCTGCCTGTAAGCGCCAGC 1080
 L G L S L N K R K G C L I A A C K R Q Q
 AAGTAACGGGATTGTTGTTAATCACAAGCCACAGCTTGCCCGTGAAGCGCGCGGGGCGC 1140
 V T G I V V N H K P Q L A R E A R R A L
 TGGCTCAGGAGGTGCATTGTGCCAAAAATATGGCGTTATTTTCGCATCTTAGTCATCGTG 1200
 R Q E V H L C Q K Y G V I S H L S H R G
 GTGAACCTTGATCCTTCTGGCGATCTCCACGCACAGGCAACGGCGTATCTTTATGCTTTGC 1260
 E L D P S G D L H A Q A T A Y L Y A L Q
 AGGGAAGAATAAACTGGTTATTGCAATCAACCTGAGGATGAGGCTTTCAACAGGCGA 1320
 G R I N W L L Q I N P E D E A F Q Q A R
 GAGAGAGTGTAAGCGAATGCTGGTTGCATGGTAAGAAAAGCGTCAGGCAGACGTTTCTG 1380
 E S V K R M L V A W *
 CCTGACCGTTTAGGGGAGAattactgcaactgcgcggaattagcggccagcggcggtca 1440
 aaatcatcgtcgggcggtatttaactcgctgcggaacaaacgtgacagcataccttca 1500
 cagaaggccagatctggttgccagcagggtttcatcgg 1540
 Oligo 2336

FIGURE 19

25E0E0-TE0B0B30

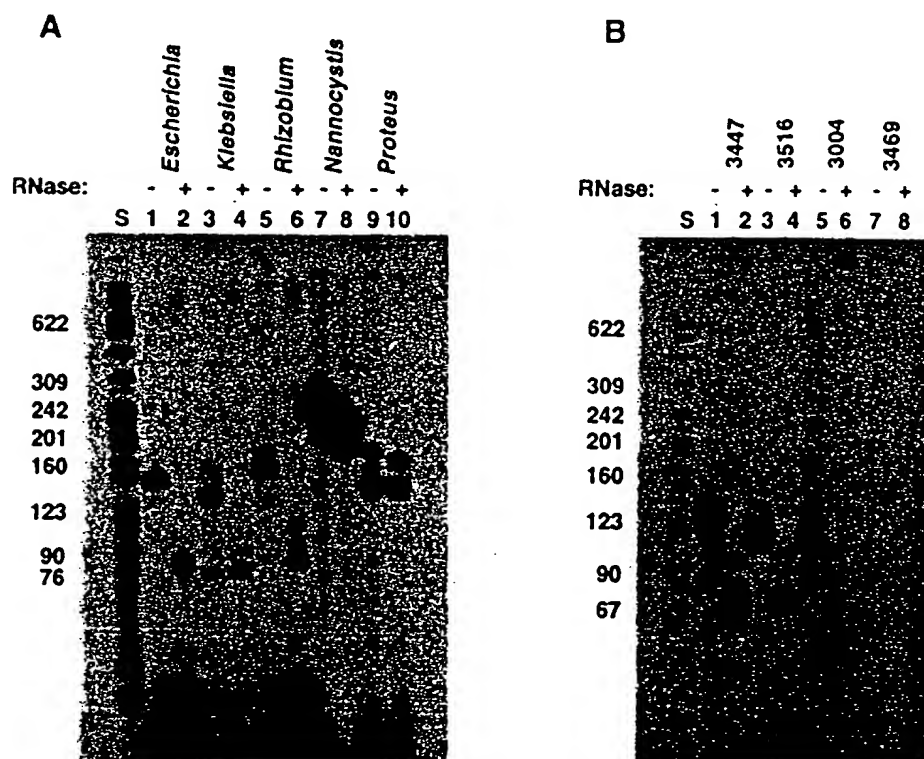


FIGURE 20

RHIZOBIAL ISOLATES

Strain (legume host genus)	USDA strain no.	Geographic source (date)	msDNA produced ^b
<i>Rhizobium</i> sp. (<i>Acacia</i>)	3002	Brazil (1959)	+
	3003	Africa (1950)	
	3325	Morocco (1974)	
	3838	? (1976)	
<i>Bradyrhizobium</i> sp. (<i>Aeschynomene</i>)	3516	Florida (1972)	+
	4362		+
<i>Bradyrhizobium</i> sp. (<i>Albizia</i>)	3004	Maryland (1952)	+
<i>Bradyrhizobium</i> sp. (<i>Apios</i>)	3240	Maryland (1939)	
<i>Bradyrhizobium</i> sp. (<i>Arachis</i>)	3339	Thailand (1979)	
	3341	Hawaii (1978)	
<i>Rhizobium</i> sp. (<i>Astragalus</i>)	3854	Alaska (1962)	
<i>Rhizobium</i> sp. (<i>Cajanus</i>)	3472		
<i>Bradyrhizobium</i> sp. (<i>Canavalia</i>)	3317	Brazil (1974)	
<i>Rhizobium</i> sp. (<i>Cicer</i>)	3378		
	3379	Mexico (1963)	
<i>Bradyrhizobium</i> sp. (<i>Coronilla</i>)	3165	Virginia (1935)	
	3167	? (1961)	
<i>Bradyrhizobium</i> sp. (<i>Crotalaria</i>)	3384	Brazil (1967)	
<i>Bradyrhizobium</i> sp. (<i>Desmodium</i>)	3225	Ecuador (1948)	
<i>Bradyrhizobium</i> sp. (<i>Erythrina</i>)	3241		
	3242	Maryland (1939)	+
<i>Rhizobium fredii</i>	191	China (1979)	
<i>Rhizobium leguminosarum</i>	2370	Illinois (1933)	
	2429	Hawaii (1978)	
	2435	Holland (1955)	
	2480	Tennessee (1951)	
	2489		
<i>Rhizobium</i> sp. (<i>Lens</i>)	2426		
	3404	Colombia (1979)	
<i>Rhizobium loti</i>	3084	Maryland (1946)	
	3468	New Zealand (1961)	+
	3469		
	3471		
	3503		
<i>Bradyrhizobium</i> sp. (<i>Lotus</i>)	3669	California (1968)	
	3074	Minnesota (1954)	
	3470	California (1916)	
<i>Rhizobium</i> sp. (<i>Lupinus</i>)	3040	Florida (1940)	
<i>Bradyrhizobium</i> sp. (<i>Lupinus</i>)	3045	Florida (1946)	
<i>Bradyrhizobium</i> sp. (<i>Macrotyloma</i>)	3451	Zimbabwe (1960)	
<i>Rhizobium medicago</i>	1097	North Dakota (1948)	
<i>Rhizobium meliloti</i>	1011	Maryland (1933)	
	1021a	North Dakota (1948)	
<i>Rhizobium phaseoli</i>	2667	Washington (1948)	
	2669		
	2674	Brazil (?)	
	2676	Colombia (1972)	
	3256	Illinois (1941)	
<i>Rhizobium</i> sp. (<i>Robinia</i>)	3436		
<i>Bradyrhizobium</i> sp. (<i>Stylosanthes</i>)	3441	Brazil (?)	
	3477	Colombia (1976)	
<i>Rhizobium trifolii</i>	2046	Virginia (1934)	
	2048	Illinois (1934)	+
	2063	Florida (1939)	
	2065	Alabama (1952)	
	2116	South Carolina (1944)	
	2134	? (1974)	
	2145		
	2156	California (1920)	
<i>Rhizobium</i> sp. (<i>Trigonella</i>)	1177	Florida (1939)	
<i>Rhizobium tropici</i>	2744	Brazil (?)	
<i>Bradyrhizobium</i> sp. (<i>Vigna</i>)	3447	Thailand (1979)	+
	3456	Wisconsin (1966)	

^a All strains are from the USDA Beltsville Rhizobium Culture Collection, provided by Peter van Berkum.

^b As defined by detection of radiolabeled msDNA by the RT extension method.